

dc_718_13

MTA Doktori Pályázat

TÉZISEK

**Transzkriptom szintű adatok alkalmazása a
rosszindulatú daganatos betegek várható
terápiás válaszána és túlélésének
előrejelzésére**

Dr. Györfly Balázs, PhD

Budapest, 2013

1. BEVEZETŐ

A rosszindulatú daganatos betegségek előfordulása világszerte nő. Az EU-28-ban 1990-hez viszonyítva másfélszeresére (Magyarországon majdnem kettő és félszeresére!) nőtt a daganatos eredetű halálozás. A betegek gyógyítása során három fő lehetőség áll rendelkezésünkre, amely a sebészi eltávolítás és sugarazás mellett a teljes test valamennyi sejtjét elérő („szisztémás”) gyógyszeres kezelést tartalmazza. A szisztémás kezelés előnye, hogy a rejtve maradó izolált tumorsejtekre is tud hatni.

A gyógyszeres kezelés legfontosabb gátja a rendelkezésre álló szerekkel szembeni rezisztencia. Általánosan a rezisztenciát több altípusra lehet osztani, mint szerzett és eredendően meglevő rezisztencia vagy egy, vagy több gyógyszerrel szembeni rezisztencia - ezen utóbbit a szakirodalomban multidrog-rezisztenciának (MDR) nevezzük. A kombinációban alkalmazott terápia fontossága miatt az MDR intenzív kutatások tárgya. Számos, a kialakulásához vezető mechanizmust azonosítottak, mint a gyógyszerek sejten belüli felhalmozódását megakadályozó pumparendszerek (pld. MDR1 gén), a programozott sejthalál (apoptózis) útvonalaiban résztvevő gének (pld. TP53) és a sejtciklus ellenőrző pontjainak változása. Egyes tumorokban ezek akár egyszerre, egymástól függetlenül is kialakulhatnak.

A gyógyszerekkel szembeni érzékenység előrejelzése megfelelő biomarkerek alkalmazásával kulcsfontosságú lesz a jövőben, mert a személyre szabott gyógyszeres terápia széleskörű elterjedésének alapja lehet. Az NIH szerint biomarkernek nevezünk egy olyan jellemzőt, amelynek objektív mérése lehetővé teszi, hogy egy normális biológiai folyamat vagy egy patológiai folyamat vagy egy terápiás célból alkalmazott farmakológiai kezelésre adott válasz indikátoraként alkalmazzuk. A „személyre szabott terápia” alkalmazása során a rendelkezésre álló lehetőségek közül az adott beteg számára legmegfelelőbb gyógyszert vagy kezelési protokollt választjuk ki.

2. CÉLKITŰZÉSEK

Vizsgálataink során onkológiai biomarkerek kutatásával foglalkoztunk. Ezen belül három fő célt tűztünk ki, úgymint:

- 2.1. a legjobb biomarker-jelöltek azonosítása transzkriptom szintű adatok alkalmazásával *in vitro* kísérletekben**
- 2.2. korábban azonosított biomarkerek igazolása transzkriptomikus adatok kereszt-elemzésével és klinikai mintákon való teszteléssel**
- 2.3. bioinformatikai eszközök fejlesztése, amelyek a transzkriptomikus adatok alkalmazását lehetővé teszik további biomarkerek vizsgálata során.**

3. MÓDSZEREK

In vitro kísérletek

A vizsgálatok során rosszindulatú daganatos sejtvonalakat használtunk, amelyekben SRB és Roche MTT (3-[4, 5-dimethylthiazolyl-2]-2, 5-diphenyltetrazolium bromid) sejtproliferációs kitek felhasználásával mértük a sejtek vitalitását. A méréseket általános ATCC vonalakon végeztük. Egyes kísérletek alapját olyan, a kutatócsoportunk tagjai által korábban létrehozott leány-sejtvonalak adták, amelyek egy-egy kiválasztott gyógyszerrel szemben rezisztens derivátumait és a szülői sejtvonal kombinációját tartalmazták. Ezek között a vonalak között az EPG85-257P, az EPP85-181P, a HT29, az MCF7, az MDA-MB-231 valamint a MeWo vonalak és leány-vonalak szerepeltek. A logaritmikus növekedés fázisában összesen 10^7 sejtől izoláltunk RNS-t Qiagen RNeasy kit segítségével. Az izolált RNS minőségét LabChip Bioanalyser-en ellenőriztük.

Párhuzamos evolúciós modellrendszer létrehozása

Egy kísérletben az MDA-MB-231 és az MCF-7 sejtvonalakból rezisztens leányvonalakat fejlesztettünk, amelynek során a szülői vonalakból 20-20 alvonalat különítettünk el, amelyeket a fejlesztés teljes ideje alatt elkülönítve kezeltünk (tehát összesen 40 sejtvonalat). A rezisztenciát két gyógyszerrel, a doxorubicinnal és a paclitaxellel szemben hoztuk létre a sejtvonalak hetente történő kezelésével.

A kromoszomális instabilitás mérésére hagyományos citogenetikai méréseket végeztünk, majd Shannon diverzitási indexet számoltunk a 3-as, 17-es és 21-es kromoszómák variációinak tíz metafázisban történő számításával. Áramlásos citometria mérést végeztünk annak megállapítására, hogy a PGP transzportrendszer milyen mértékben változott meg az egyes sejtvonalakban, ennek során a sejtszámot a rhodamin intenzitáshoz viszonyítva ábrázoltuk.

Génelcsendesítés

Az RNS alapú génelcsendesítés („RNA interference”) elvégzéséhez szükséges rövid oligonuklotid szakaszokat Applied Biosystems Silencer siRNA Construction Kit felhasználásával készítettük. Az siRNS kezeléshez a transzfekciót Applied Biosystems SiPort NeoFX transzfekciós reagens felhasználásával, 230 ezer sejten végeztük. Az siRNS hatásosságát RT-PCR alkalmazásával igazoltuk, a PCR mérésekhez a primereket a Primer3 programmal terveztük. A PSMB7 gén hatását a kemorezisztenciára siRNS és gyógyszeres kezelés kombinációjával vizsgáltuk.

Transzkriptomikus mérések

A transzkriptomikus adatokat Affymetrix HGU95 és HGU133A valamint saját tervezésű, 43 ezer próba mérésére alkalmas gén chipek felhasználásával mértük meg, amelyeket a berlini Charité Funkcionális Genomikai Központjában, a Stanford-i Egyetemen és a Semmelweis Egyetemen végeztünk el. Egyes kiválasztott gének expresszióját saját készítésű RT-PCR-el, illetve TaqMan kvantitatív géneszpressziós méréssel igazoltuk.

Klinikai mintagyűjtés

A Semmelweis Egyetem Urológiai Klinikáján 2005 és 2010 között folytattunk veserák mintagyűjtést, a Semmelweis Egyetem I. Nőgyógyászati Klinikáján és az Országos Onkológiai Intézetben pedig petefészekrák mintákat gyűjtöttünk 2005 és 2008 között. A betegekből a mintagyűjtés a gyógyszeres kezelés előtt történt, ami a sejtkultúrás kísérletek során végzett megközelítéssel egyezett.

Bioinformatikai kiértékelések

A statisztikai számítások az R környezetben (<http://www.R-project.org>) történtek, Bioconductor csomagok felhasználásával (<http://www.bioconductor.org>). Az adatok normalizálására a MAS5, az RMA és a VSN algoritmusokat használtuk. A tanulóhalmazban a jellemzők kiválasztására „prediction analysis of microarrays” (PAM), a „significance analysis of microarrays” (SAM) és a „rank products” algoritmusokat használtuk. Túlélési számítások esetén Cox regressziót számoltunk és Kaplan-Meier túlélési görbéket készítettünk. Fürtelemzést a Genesis program felhasználásával végeztünk. További elemzéseket a BRB ArrayTools felhasználásával végeztünk. Az RT-PCR méréseket az SDS 2.0 program használatával értékeltük ki. Multivariancia elemzést Winstat 2007 alkalmazásával végeztünk.

Adatbázis építés publikus transzkriptomikus adatokból

A PubMed GEO-ban és az EGA-ban keresést végeztünk a „tumortípus”, „rák”, „GPL96”, „GPL570” és a „GPL571” kulcsszavak felhasználásával. A keresések során a tumortípus emlőrák, petefészekrák és melanoma voltak. A rendszer több Affymetrix platform egyesítését tartalmazta, mint a GPL96-os (HGU133A), a GPL570-es (HGU133 Plus 2.0) és később a GPL571-es (HGU133A 2.0) gén chippek. Ezen platformok tartalmazznak 22277 közös próbát, amelyekben az egyes gének mérésére használt próbák szekvenciája egymással teljesen megegyezik, ezért a hibridizáció során azonos érzékenységet és dinamikus tartományt várhatunk.

Online rendszerek fejlesztése

A rendszer három részből áll, úgymint az adatok egy PostgreSQL adatbázisban, a világhálón keresztül elérhető kezelőfelület és az R statisztikai környezetet futtató virtuális kiszolgáló. Az R környezetet futtató virtuális kiszolgálóra („R server”) azért volt szükség, mert az R-ben elérhető könyvtárak („libraries”) biztosították az eszközöket a statisztikai számítások elvégzéséhez. Azonban az R alap esetben önálló program, amelyet a felhasználó parancssorból tud kezelni. A rendszerben a fizikális kiszolgálón („server gépen”) elindított virtuális R kiszolgáló megkapja az adatokat az adatbázis kezelőből, majd a számítások elvégzése után visszaadja az eredményt. A kiszolgáló operációs rendszere Debian Linux (www.debian.org), amely egy Apache webkiszolgálót (www.apache.org) futtat. Az R és a PostgreSQL adatbázis között a kapcsolatot az RODBC csomag biztosítja (<http://cran.r-project.org/package=RODBC>). A kezelőfelületet PHP-ban készítettük el. Az online diagnosztikus rendszer honlapját a Drupal (www.drupal.org) tartalomkezelő környezetben hoztuk létre.

4. EREDMÉNYEK

4.1. A legjobb biomarker-jelöltek azonosítása genom-szintű adatok alkalmazásával *in vitro* kísérletekben

4.1.1. Kemoterápiás szerekkel szembeni rezisztencia-profil meghatározása sejtvonalak vizsgálatával

Ezen első, transzkriptom szintű adatot is alkalmazó vizsgálatunk során célunk az volt, hogy az egyes tumortípusoktól függetlenül tudjunk kemorezisztencia-biomarkereket azonosítani. Figyelembe véve, hogy az elképzelések alapján a tumorok általában hasonló rezisztencia-mechanizmusokat alakíthatnak ki, ezért arra törekedtünk, hogy nagyszámú sejtvonal egyidejű vizsgálatával olyan markereket azonosítsunk, amelyek a rosszindulatú daganatos betegség típusától függetlenek. A kutatás során összesen 30 különböző sejtvonal génexpressziós mintázatának felhasználásával 11 rákellenes gyógyszerrel szembeni rezisztenciagéneket határoztunk meg. Eredményeink alapján:

- Sejtvonalak felhasználásával sikeresen tudtunk olyan géneket és génexpressziós mintázatokat azonosítani, amelyek a gyógyszerekkel szembeni rezisztenciát előre tudják jelezni.
- Sikerült azonosítanunk a gének egy olyan csoportját (pld TRIM2, APOBEC3B, BIRC2), amelyek több gyógyszerrel szembeni rezisztenciában is fontos szerepet bírhatnak.
- A DNS microarray technológia alkalmas arra, hogy az érzékenység és rezisztencia transzkriptom szintű markereit azonosítani tudjuk.

4.1.2. A kemorezisztencia génjei melanoma malignumban

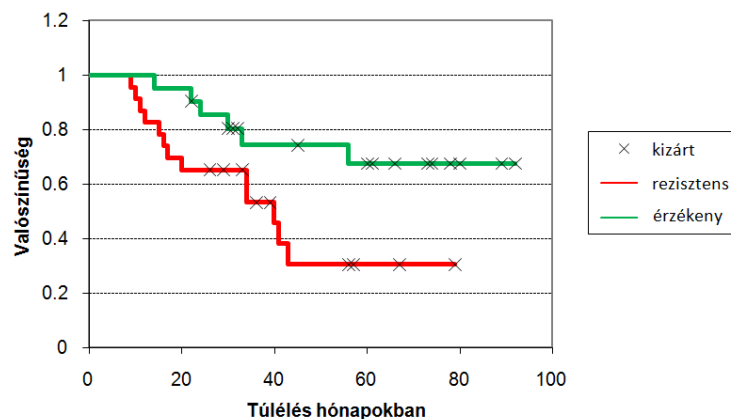
Rendelkezésünkre állt egy szülői melanoma sejtvonal (MeWo), valamint ennek négy különböző gyógyszerrel szemben rezisztenssé tett leányvonalai. A vizsgálatok során a szülői- és leány-vonalak génexpressziós mintázatát határoztuk meg, hogy ezáltal olyan géneket azonosítsunk, amelyek az egyes gyógyszerekkel szembeni rezisztenciával kapcsolatba hozhatóak. A kutatásunk eredményei alapján a következőket állapítottuk meg:

- Rezisztens és érzékeny sejtvonalpárok felhasználásával lehetséges volt olyan géneket azonosítani, amelyek különbözőképpen fejeződnek ki az egyes vonalakban.
- Fürtelemzés segítségével az egyes vonalak közötti hasonlóságot tudtuk azonosítani.
- Az irodalomban korábban leírt géneket (pld STK17A, CPNAS) sikeresen tudtuk igazolni gén chip mérések felhasználásával.

4.1.3. Várható túlélés előrejelzése sejtkultúra-modellben azonosított gének felhasználásával

Kutatásaink során a sejtkultúrák modell rendszerben kemoterápiával szemben ellenálló és érzékeny sejtvonalak génexpressziós mintázatát hasonlítottuk össze, majd az azonosított mintázattal klinikai betegekben előrejelzést adtunk a várható terápiás választ illetően. Összesen négy különböző tumortípust vizsgáltunk egyidejűleg, hogy az egyes szövettípusoktól függetlenül csak a kemoterápiás válasszal kapcsolatba hozható géneket tudjuk azonosítani. Végezetül az adott gyógyszer (doxorubicin) monoterápiájával kezelt betegek elemzésével teszteltük modellrendszerünk hatásosságát. Eredményeink alapján:

- Rezisztens és érzékeny sejtvonal-párok felhasználásával két gyógyszerrel (doxorubicin és paclitaxel) szemben azonosítottuk a rezisztencia legfontosabb génjeit.
- Egy független, 44 beteget tartalmazó betegmintán alkalmazva a génexpressziós mintázatuk alapján általunk doxorubicin kemoterápiára érzékenyként osztályozott betegek túlélése több mint 50%-al hosszabb volt, mint a rezisztensként osztályozott betegek túlélése.
- Vizsgálatunk igazolta az *in vitro* rendszerek használhatóságát klinikailag is alkalmazható előrejelző tesztek kifejlesztésére.



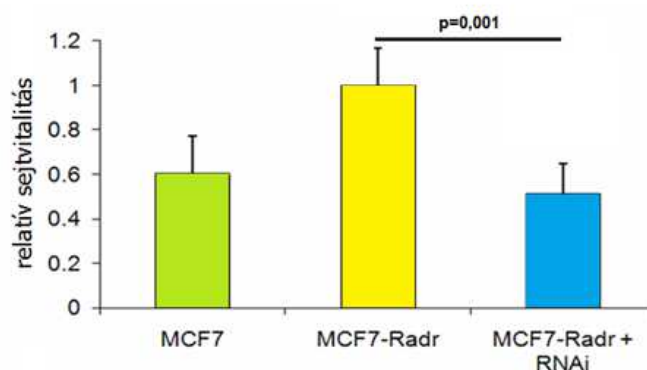
1. ábra. Doxorubicin monoterápiában részesült betegek túlélése gén chip adatok alapján rezisztensként és érzékenyként osztályozott betegek között szignifikánsan eltért ($p=0,0334$). Az osztályozás a rezisztens, illetve érzékeny sejtvonalakban mért génexpressziós mintázathoz való hasonlóság alapján történt.

4.1.4. Transzkriptomikus adatokkal meghatározott gének oki szerepének igazolása kísérletes körülmények között

A kutatás során először gén chip technológia alkalmazásával a doxorubicin-rezisztenciával kapcsolatba hozható proteaszóma-alegységeket azonosítottuk. A projekt második fázisában ezen alegységek rezisztenciában betöltött szerepét kísérletes módon igazoltuk, amelynek során a PSMB7 gén elcsendesítésével egyidejűleg alkalmazott gyógyszeres kezelés mellett mértük az életben maradt sejtek

arányát. Végül 1592 betegben vizsgáltuk a génexpresszió és a túlélés közötti kapcsolatot. Célunk volt, hogy a rezisztencia egy olyan oki génjét azonosítsuk, amely akár gyógyszerekkel is befolyásolható lehet, és ezáltal egy kombinált kemoterápiás kezelés alapját adhatja. A vizsgálat eredményei alapján a következő megállapításokat tettük:

- A kezelés és a rezisztencia közötti genom-szintű különbségeket felhasználva azonosítottuk a proteasóma fehérjekomplex PSMB7 alegységét, mint a doxorubicinnel szembeni rezisztencia egyik oki („driver”) génjét.
- A vizsgálat során igazoltuk, hogy a genom-szintű adatokkal azonosított célgének kísérletes körülmények között is magas relevanciával rendelkeznek.
- A nagy klinikai mintán is igazolt eredmények alapján a rossz prognózissal kapcsolatba hozható PSMB7 gén egy ígéretes terápiás célpont lehet.



2. ábra. A PSMB7 csendesítésének hatása a sejtek relatív vitalitására gyógyszeres kezelés mellett a kezeletlen vonalakhoz viszonyítva doxorubicin kezelés során. Az érzékeny szülői vonal (zöld) nagymértékű pusztulást mutat, a rezisztens leány-vonal (MCF-7-RAdr, sárga) vitalitás-csökkenése nem észlelhető. Az elcsendesített PSMB7 mellett a sejtek vitalitása az érzékeny szülői vonalakéval volt azonos.

4.1.5. Parallel evolúciós modell alkalmazása a párhuzamosan létrejövő kemorezisztencia modellezésére

A fenti vizsgálatok során alkalmazott sejt kultúrák modellek alapja az a feltételezés, hogy a kísérletes körülmények között tartott sejtek a való világhoz hasonló rezisztenciamechanizmusokat fognak „alkalmazni”. Ezen hasonlóság vizsgálatára párhuzamosan hoztunk létre rezisztens vonalakat, majd ezeket egymással hasonlítottuk össze. Az eredményeink alapján:

- A sejt vonalakban történő párhuzamos kezelés révén sikerült 27 sejt vonalban létrehozni a párhuzamos evolúció modelljét egy konkrét szelekciós nyomással szemben.
- Több, korábban már leírt különféle mechanizmus aktiválódását (pld TOP2A, TUBB3, ABCB1) tudtuk igazolni a rezisztencia létrejötté során.

- Mindössze 1-2 mechanizmus aktiválódása már elégséges volt ahhoz, hogy a rezisztencia létrejöjjön.

4.1.6. Célzott terápiás szerekkel szembeni rezisztencia és érzékenység vizsgálata

Egy újabb vizsgálatban az egyes jelátviteli utak modulálásán keresztül ható öt célzott terápiás gyógyszerrel (lapatinib, erlotinib, sorafenib, sunitinib és gefitinib) szembeni rezisztencia markereit azonosítottuk 45 sejtvonal felhasználásával *in vitro*. A legjobb sunitinib biomarker-jelölt géneket 39 független klinikai mintán is igazoltuk. Eredményeink alapján:

- Az alkalmazott, sejtkultúra és transzkriptom szintű elemzések kombinációjára épülő megközelítés alkalmasnak bizonyult arra, hogy a célzott terápiás szerekkel szembeni relatív rezisztenciát előrejelző biomarkereket azonosítsunk.
- Veserákos mintákban igazoltuk a RAB17, az LGALS8 és az EpCAM gének összefüggését a sunitinib-rezisztenciával.
- A legtöbb keresztrezisztenciával összefüggő gént a sunitinib-rezisztenciával kapcsolatba hozott génlista tartalmazta.

4.2. Korábban azonosított biomarkerek igazolása transzkriptom szintű adatok keresztelemzésével és klinikai mintákon való teszteléssel

4.2.1. Korábbi közleményekben a melanoma pathogenezisével kapcsolatba hozott génlisták összehasonlítása

Az 4.1.2. részben már bemutattuk, hogy saját sejtkultúra alapú vizsgálataink során tudtunk a melanoma kemorezisztenciájával összefüggő géneket azonosítani. Kérdésként vetődik fel, hogy az irodalomban azonosított génexpressziós mintázatok mennyire reprodukálhatóak független betegmintákban is. A vizsgálat során az elérhető transzkriptomikus adatok alapján azonosított biomarkereket hasonlítottuk össze és a hozzájuk rendelhető gén-ontológiai kategóriákat határoztuk meg egy keresztelemzés során. A kutatás feltételezése volt, hogy nem szükséges az alapadatokat minden esetben ismernünk, elég a kutatások végeredményét elemezni. Eredményeink alapján:

- A melanoma génexpressziós mintázatok alapján generált génlisták különbözősége a tumorok nagyfokú heterogenitását mutatja.
- A génexpressziós mintázatok generáló experimentális és klinikai vizsgálatok reprodukálhatósága alacsony.

4.2.2. Korábbi közleményekben emlőrák prognózisával kapcsolatba hozott génlisták kereszt-elemzése nyers adatok újrafeldolgozása után

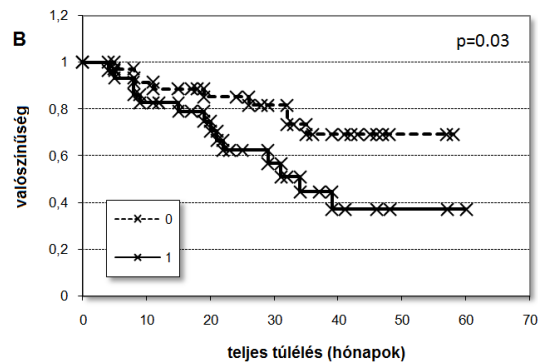
Az előző fejezetben tárgyalt, általunk végzett első kereszt-elemzési vizsgálat egy jelentős gyengesége, hogy a nyers adatok hiányában az eredeti szerzők által már feldolgozott adatok alapján készített génlistákat hasonlítottuk egymással össze. Ezután eredeti közlemények nyers adatait kombináltuk és teljesen újra feldolgoztuk, majd ebben az integrált adatbázisban ellenőriztük az egyes gének és génexpressziós mintázatok valódi teljesítményét. Saját eredményeink alapján:

- Korábban közölt adatok felhasználásával létrehoztunk egy integrált, klinikai és génexpressziós adatokat tartalmazó adatbázist, amely a prognosztikus biomarkerek kereszt-elemzését lehetővé teszi.
- Az irodalomban közölt mintázatok többsége nem volt képes a jó és rossz prognózisú betegeket egymástól elkülöníteni. Az általunk szignifikánsnak talált génlisták (n=8) hatásosságát eközben egy független platformon igazoltuk.
- Az adatbázis felhasználásával azonosított gének közül kevesebb is elég egy új osztályozóba, ha a klinikai paramétereket is figyelembe vesszük.
- Nagyszámú redundáns gén chipet közöltek korábbi közleményekben (n=641).

4.2.3. Petefészekrákok szövettani altípusait meghatározó és a várható túlélést előrejelző gének azonosítása és igazolása klinikai mintákban

A következő kutatásunk során először nagy *in silico* adatbázist építettünk transzkriptomikus adatokkal (n=829), amelyen a szövettani altípusokkal és a túléléssel összefüggő géneket azonosítottuk. Eközben klinikai mintagyűjtést folytattunk a Semmelweis Egyetemen és az Országos Onkológiai Intézetben. A gén chipek segítségével azonosított géneket 64 klinikai mintában egy független technológia (RT-PCR) alkalmazásával igazoltuk. Eredményeink alapján:

- Az általunk elvégzett kiértékelés során azonosított, a túléléssel legjobban összefüggő gének a PGR (lásd ábra), az ÖR2 és a TSPAN8 voltak.
- Az általunk gyűjtött klinikai mintákban RT-PCR alkalmazásával a gén chipeken azonosított gének több mint 90%-át sikeresen tudtuk igazolni.
- A korábban közölt génlisták közül nyolc volt képes a szövettani altípus elkülönítésére, de egy sem tudta a túlélést előre jelezni.



3. ábra. Kaplan-Meier elemzés a PGR génre 64 petefészekrákos betegben. A gén expressziós értékeinek mediánja alapján osztottuk két csoportra a betegeket. (1: magas, 0: alacsony expressziót mutató minták)

4.3. Bioinformatikai eszközök fejlesztése, amelyek a transzkriptomikus adatok alkalmazását lehetővé teszik további biomarkerek vizsgálata során

4.3.1. Microarray adatok előfeldolgozása

A korábbi kutatásaink során a mérési adatokat gén chipek (más néven génexpressziós microarray-ek) felhasználásával generáltuk. A technológia alkalmazása során az eredményt egy leolvasó („scanner”) segítségével határoztuk meg, amely egy kép fájl formájában menti le a mért adatokat. Ezen képfájlból az egyes képpontokat elhelyezkedés alapján azonosítja a program, hozzárendeli az adott pozícióban elhelyezett próbához, majd a kapott intenzitási értékeket egy következő fájlban tárolja el. Az egyes génekre vonatkozó kifejeződést ezen fájl további feldolgozásával tudjuk kiszámolni, amely folyamatot előfeldolgozásnak („pre-processing”) nevezzük, mivel ezt még az egy kísérleten belül elvégzett több mérés összehasonlítása előtt le kell futtatni. Figyelembe véve, hogy kutatásaink során egyrészt mi magunk is jelentős számú microarray mérést végeztünk, másrészt a különböző vizsgálatok keresztlemzése, illetve a később tárgyalt további fejlesztések előfeltétele az adatok összehasonlíthatósága, ezért szükséges volt az elérhető kilenc előfeldolgozási algoritmus részletes elemzése. Az algoritmusokat általunk sejtvonalakból és klinikai mintákból készített gén chip és RT-PCR adatok összehasonlítása során vizsgáltuk. Számításaink alapján:

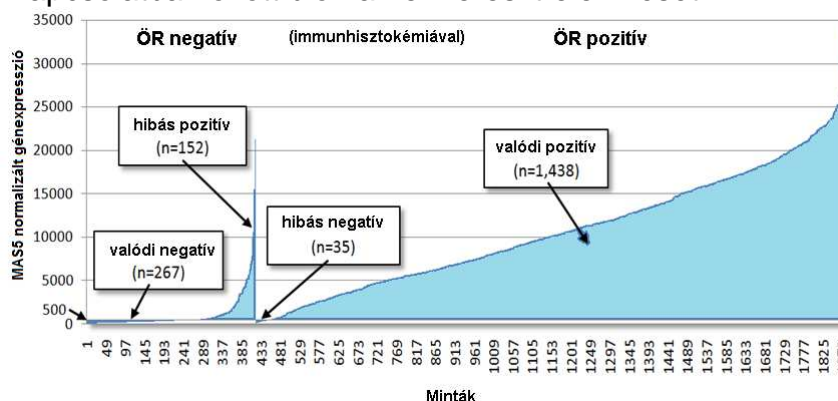
- a legjobban teljesítő előfeldolgozási algoritmus a PLIER+16 volt, és
- a legelterjedtebb algoritmusok hasonlóan jól teljesítettek.

A kereszt-elemzések és rendszerfejlesztések során célunk volt az adatbázisok folyamatos bővítése és továbbfejlesztése, ezért a jól teljesítő algoritmusok között a MAS5-öt választottuk, amely egyetlen gén chipre is elvégezhető.

4.3.2. Online rendszer fejlesztése az emlőrák prognózisát előrejelző génexpresszió alapú biomarkerek tesztelésére

A kutatás során a rendelkezésünkre álló gén chip adatokra építve egy olyan online elérhető rendszert fejlesztettünk ki, ahol 22277 gén expressziójának az egyes betegekben mért szintje és a betegek várható kiújulás mentes és teljes túlélése közötti kapcsolatot lehet vizsgálni. A rendszerbe több ezer beteget integráltunk, amely lehetővé teszi, hogy egyes ritkább alcsoportokban is megbízható statisztikai számításokat lehessen végezni. A program létrehozása után a korábban leírt biomarker jelöltek közül a legígéretesebbek vizsgálatával demonstráltuk a software alkalmazhatóságát. Bioinformatikai fejlesztéseink alapján a következő megállapításokat tettük:

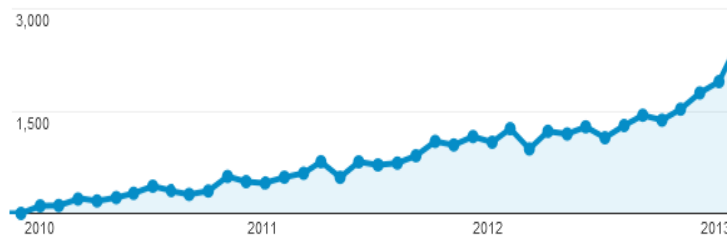
- Létrehoztunk egy, a világhálón keresztül elérhető rendszert (www.kmplot.com), amely lehetővé teszi a rendelkezésre álló transzkriptomikus adatok kereszt-elemzésével 22277 gén biomarker-szerepének valós időben történő vizsgálatát.
- A lekérdező felületet kiegészítettük szűrőkkel, amelyek alkalmazásával az elemzést egy-egy kiválasztott klinikai csoportban (*pld* ösztrogén-receptor pozitív és negatív betegek, lásd ábra) is el lehet végezni.
- A rendszer felhasználásával elvégeztük a korábban leírt legfontosabb 11, túléléssel kapcsolatba hozott biomarker kereszt-elemzését.



4. ábra. A gén chip alapú és az immunhisztokémia alapú ösztrogén státusz kapcsolata 1984 beteg adatai alapján. A vágópontként alkalmazott 500-as értéket az Y tengelyen nyíl jelöli. Az chip alapú ÖR státusz meghatározás eredményét ezután a 4.3.2. pontban leírt online rendszerbe szűrőként építettük be.

4.3.3. A világhálón keresztül elérhető kereszt-elemző rendszer továbbfejlesztése többgénes osztályozók és optimalizált vágópont számítására

Az előző pontban említett Kaplan-Meier rajzoló programot továbbfejlesztettük, hogy több gén egyidejű mérésére alkalmas legyen és lehessen a vágópontot optimalizálni. A fejlesztés során petefészekrák mintákat használtunk fel, és a fentebb leírt rendszerrel párhuzamosan létrehoztunk egy második elemző oldalt. A rendszer felhasználásával elvégeztük 37, az irodalomban leírt biomarker kereszt-elemzését.

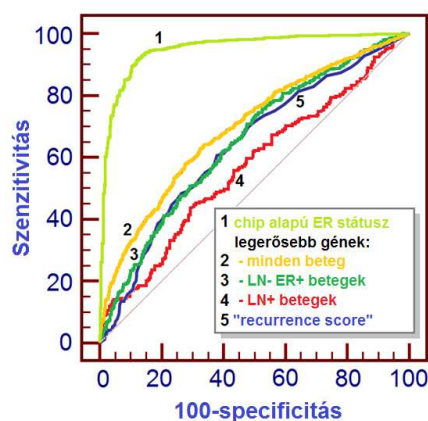


5. ábra. A adatbázis folyamatos növelésének és az új fejlesztéseknek köszönhetően a Kaplan-Meier rajzoló felhasználtsága folyamatosan növekedett az elmúlt három év folyamán. A legutolsó mért hónapban (2013 május) az egyedi látogatók száma a Google Analytics adatai alapján nagyobb volt, mint az első év folyamán összesen.

4.3.4. A világhálón keresztül elérhető diagnosztikai rendszer fejlesztése

Az eddig bemutatott bioinformatikai számítások és fejlesztések során a biomarkerek azonosítása volt a cél. A biomarkerek alkalmazására elvégzett bioinformatikai projektünk során egy, a világhálón keresztül elérhető diagnosztikai rendszert fejlesztettünk ki, amely a betegből eltávolított tumorszövet feldolgozásával készült gén chipek segítségével direkt betegdiagnózis készítésére alkalmas. A rendszerbe a három legerterjedtebb vizsgálatot (ösztrogén receptor, HER2 státusz, 21 génes „recurrence score”) építettük be, majd nagy mintaszámon (n=2472) ellenőriztük a hatásosságot. Eredményeink alapján:

- A betegek osztályozása gén chip mérések alapján lehetséges és a korábbi diagnosztikai tesztek hatásosságánál jobb.
- A létrehozott, a világhálón a www.recurrenceonline.com címen elérhető rendszer rendkívül rövid elemzési idő mellett (10 perc) lényegében 100%-os reprodukálhatóságot kínál.
- Több különböző diagnosztikai mérés kombinálása egy rendszerbe technikailag megvalósítható.



6. ábra. Az egyes osztályozók relatív teljesítményének összehasonlítása ROC elemzéssel. A várható túlélést előrejelző osztályozók esetében az ötéves túlélésig bekövetkezett eseményt használtuk fel végpontként.

A SAJÁT KUTATÁSI EREDMÉNYEK ÖSSZEFOGLALÁSA

1. A kutatások első fő csoportjában a legjobb biomarkerjelölteket azonosítottuk *in vitro* kísérletekben transzkriptom szintű adatok alkalmazásával.

- 1.1. Tizenegy kemoterápiás szerrel szembeni rezisztencia-profilt határoztunk meg 30 rosszindulatú daganatos sejtvonalban, majd gén chip adatok felhasználásával sikeresen tudtunk olyan géneket és génexpressziós mintázatokat azonosítani, amelyek a gyógyszerekkel szembeni rezisztenciát előre tudják jelezni. Azonosítottunk 22 olyan gént, amelyek a multidrog-rezisztenciában is fontos szereppel bírhatnak.
- 1.2. A kemorezisztencia génjeit melanoma malignumban rezisztens sejtvonal-párok felhasználásával azonosítottuk. Fürtelemzés segítségével az egyes vonalak közötti hasonlóságot tudtuk azonosítani, valamint igazoltuk, hogy az irodalomban korábban leírt géneket eredményesen lehet igazolni gén chip mérések felhasználásával.
- 1.3. Rezisztens és érzékeny emlőráksejtvonal párok felhasználásával két gyógyszerrel szemben azonosítottuk a rezisztencia legfontosabb génjeit. Egy független betegmintán alkalmazva a génexpressziós mintázatuk alapján általunk kemoterápiára érzékenyként osztályozott, doxorubicin monoterápiával kezelt betegek túlélése több mint 50%-al hosszabb volt, mint a rezisztensként osztályozott betegek túlélése.
- 1.4. A transzkriptomikus adatokkal azonosított proteaszóma fehérjekomplex egyik alegységének oki szerepét igazoltuk kísérletes körülmények között RNS interferencia alkalmazásával. A kezelés és a rezisztencia közötti genom-szintű különbségeket felhasználva azonosítottuk és klinikai mintákban is igazoltuk a PSMB7 gént, mint a doxorubicinnel szembeni rezisztencia egyik oki („driver”) génjét.
- 1.5. A sejtvonalakban történő párhuzamos doxorubicin és paclitaxel kezelés révén sikerült létrehoznunk a párhuzamos evolúció modelljét egy konkrét szelekciós nyomással szemben. A 27 sejtvonalban több, korábban leírt mechanizmus aktiválódását tudtuk igazolni a rezisztencia létrejötte során (pld TOP2A és ABCB1) és bizonyítottuk, hogy mindössze 1-2 mechanizmus aktiválódása már elégséges volt ahhoz, hogy a rezisztencia létrejöjjön.
- 1.6. Öt célzott terápiás szerrel szembeni érzékenységet vizsgálatuk 45 sejtvonalban és a relatív rezisztenciát előrejelző biomarkereket azonosítottuk. Vese- és májrákos mintákban igazoltuk az LGALS8, a RAB17 és az EPCAM gének összefüggését a sunitinib-rezisztenciával.

2. A következő kutatások során a cél a korábban azonosított biomarkerek igazolása volt transzkriptom szintű adatok keresztelemzésével és klinikai mintákon való teszteléssel.

- 2.1. A korábbi közleményekben a melanoma pathogenezisével kapcsolatba hozott génlisták összehasonlítása által igazoltuk nagyfokú heterogenitását génexpressziós szinten.
- 2.2. Korábban közölt emlőrákos adatok felhasználásával létrehoztunk egy integrált, 1079 beteg klinikai és génexpressziós adatait tartalmazó adatbázist, amely korábbi prognosztikus biomarkerek kereszt-elemzését lehetővé teszi.
- 2.3. Petefészekrákok szövettani altípusait meghatározó és a várható túlélést előrejelző gének azonosítására építettünk fel adatbázist 829 betegminta felhasználásával. Az általunk készített gén chip adatbázis kiértékelésével azonosított, a túléléssel legjobban összefüggő gének a PGR, az ÖR2 és a TSPAN8 voltak, amelyek szerepét 64 saját betegben is igazoltuk.

3. Kutatásaink harmadik fő csoportjában expressziós gén chip adatok feldolgozását lehetővé tévő bioinformatikai rendszereket fejlesztettünk.

- 3.1. Először a gén chip adatok előfeldolgozására elérhető algoritmusokat hasonlítottuk össze. A legerterjedtebb algoritmusok (MAS5, PLIER, RMA) hasonlóan jól teljesítettek.
- 3.2. Létrehoztunk egy, a világhálón keresztül elérhető rendszert, amely lehetővé teszi 2977 emlőrákos minta transzkriptomikus adatának kereszt-elemzésével 22277 gén biomarker-szerepének valós időben történő vizsgálatát. A rendszer felhasználásával elvégeztük a korábban leírt 11 legfontosabb biomarker kereszt-elemzését.
- 3.3. Ezt a rendszert elkészítettük petefészekrák biomarker-jelöltek tesztelésére is. A program felhasználásával elvégeztük 37, az irodalomban leírt biomarker kereszt-elemzését.
- 3.4. A biomarkerek alkalmazására elkészítettünk egy világhálón keresztül elérhető diagnosztikai rendszert, amely emlőrákos betegek várható túlélését tudja előre jelezni, valamint ösztrogén- és HER2 receptor státusz meghatározására is alkalmas. A betegek osztályozása gén chip mérések alapján a korábbi diagnosztikai tesztekhez képest magasabb hatásosságot és reprodukálhatóságot kínál.

AZ ÉRTEKEZÉS ALAPJÁUL SZOLGÁLÓ KÖZLEMÉNYEK

Közlemény	IF
1. Pénczvártó Zs, Tegze B, Szász AM, Sztupinszki Zs, Likó I, Szendrői A, Schäfer R, Györfly B. Identifying Resistance Mechanisms against five Tyrosine Kinase Inhibitors Targeting the ERBB/RAS Pathway in 45 Cancer Cell Lines. PLoS One , 2013;8(3):e59503. doi: 10.1371/journal.pone.0059503. Epub 2013 Mar 29.	3.730*
2. Magnani L, Stoeck A, Zhang X, Lánczky A, Mirabella AC, Wang TL, Györfly B, Lupien M. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. Proc Natl Acad Sci U S A . 2013 Apr 16;110(16):E1490-9. doi: 10.1073/pnas.1219992110. Epub 2013 Apr 1. PubMed PMID: 23576735.	9.737*
3. Györfly B, Lánczky A, Szállási Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data of 1287 patients, Endocrine-Related Cancer . 2012 Apr 10;19(2):197-208.	5.261
4. Györfly B, Benke Z, Lánczky A, Balázs B, Szállási Z, Timár J, Schäfer R. RecurrenceOnline: an online analysis tool to determine breast cancer recurrence and hormone receptor status using microarray data, Breast Cancer Res Treat , 2012;132:1025–1034.	4.469
5. Fekete T, Rásó E, Pete I, Tegze B, Liko I, Munkácsy G, Sipos N, Rigó J, Györfly B. Meta-analysis of gene expression profiles associated with histological classification and survival in 829 ovarian cancer samples. Int J Cancer , 2012 Jul 1;131(1):95-105. doi: 10.1002/ijc.26364.	6.198
6. Tegze B, Szállási Z, Haltrich I, Pénczvártó Z, Tóth Z, Likó I, Györfly B. Parallel Evolution under Chemotherapy Pressure in 29 Breast Cancer Cell Lines Results in Dissimilar Mechanisms of Resistance. PLoS One . 2012;7(2):e30804. Epub 2012 Feb 2.	3.730
7. Li Q, Birkbak NJ, Györfly B, Szállási Z, Eklund AC. Jetset: selecting the optimal microarray probe set to represent a gene, BMC Bioinformatics , 2011 Dec 15;12:474.	2.751
8. Györfly B, Lánczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szállási Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1809 patients, Breast Cancer Res Treatment , 2010 Oct;123(3):725-31.	4.859
9. Györfly B, Schäfer R. Biomarkers Downstream of RAS: A Search for Robust Transcriptional Targets. Curr Cancer Drug Targets . 2010 Dec 1;10(8):858-68.	4.771
10. Munkácsy G, Abdul-Ghani R, Mihály Z, Tegze B, Tchernitsa O, Surowiak P, Schäfer R, Györfly B. PSMB7 is associated with anthracycline resistance and is a prognostic biomarker in breast cancer, Br J Cancer . 2010 Jan 19;102(2):361-8.	4.831
11. Timár J, Györfly B, Rásó E. Gene signature of the metastatic potential of cutaneous melanoma: too much for too little? Clin Exp Metastasis . 2010 Aug;27(6):371-87.	4.113
12. Györfly B, Schafer R. Meta-analysis of gene expression profiles related to relapse-free survival in 1079 breast cancer patients, Breast Cancer Res Treatment , 2009 Dec;118(3):433-41.	4.696
13. Denkert C, Budczies J, Darb-Esfahani S, Györfly B, Sehouli J, Könsgen D, Zeillinger R, Weichert W, Noske A, Buckendahl AC, Müller BM, Dietel M, Lage H. A prognostic gene expression index in ovarian cancer-validation across different independent data sets. J Pathol . 2009 Jun;218(2):273-80.	6.466
14. Györfly B, Molnár B, Lage H, Szállási Z, Eklund A. Evaluation of Microarray	4.351

Preprocessing Algorithms Based on Concordance with RT-PCR in Clinical Samples, **PLoS One**, 2009 May 21;4(5):e5645.

15. Györfly B, Dietel M, Fekete T, Lage H. A snap shot of microarray-generated gene expression signatures associated with ovarian carcinoma, **Int J Gynecol Cancer**. 2008 Nov-Dec;18(6):1215-33. 1.932
16. Galamb O, Györfly B, Sipos F, Spisák S, Németh AM, Miheller P, Tulassay Z, Dinya E, Molnár B. Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature **Dis Markers**. 2008;25(1):1-16. 2.303
17. Györfly B, Lage H. A Web-Based Data Warehouse on Gene Expression in Human Malignant Melanoma. **J Invest Dermatol**. 2007 Feb;127(2):394-9. 4.829
18. Györfly B, Serra V, Schäfer R, Dietel M, Schadendorf D, Lage H. Analysis of gene expression profiles in melanoma cells with acquired resistance against antineoplastic drugs. **Melanoma Res**. 2006 Apr;16(2):147-155. 1.704
19. Györfly B, Surowiak P, Kiesslich O, Denkert C, Schäfer R, Dietel M, Lage H. Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. **Int J Cancer**. 2006 Apr 1;118(7):1699-712. 4.693
20. Abdul-Ghani R, Serra V, Györfly B, Jurchott K, Solf A, Dietel M, Schäfer R. The PI3K inhibitor LY294002 blocks drug export from resistant colon carcinoma cells overexpressing MRP1. **Oncogene**. 2006 Mar 16;25(12):1743-52. 6.582
21. Györfly B, Serra V, Jürchott K, Abdul-Ghani R, Garber M, Stein U, Petersen I, Lage H, Dietel M, Schäfer R: Prediction of Doxorubicin Sensitivity in Breast Tumors Based on Gene Expression Profiles of Drug Resistant Cell Lines Correlates with Patient Survival, **Oncogene**. 2005;24(51):7542-51. 6.872

Az értekezés alapjául szolgáló közlemények összesen	98,878
--	---------------

ÖSSZESÍTETT TUDOMÁNYMETRIAI ADATOK

Közlemények	IF
Az értekezés alapjául szolgáló közlemények összesen	98,88
Az értekezéstől független közlemények összesen	96,08
Közlemények a PhD fokozat megszerzéséig	13,32
Összesített impakt faktor	208,3
Ebből első/utolsó szerzős közlemények	88,5
Idézettség (teljes/független)	1008 / 854
Hirsch-index	20

KÖSZÖNETNYILVÁNÍTÁS

A leírt kutatások során sokmindenkitől kaptam támogató kezet. Mivel ezeket nem lehet rangsorolni, ezért időrendben szeretném áttekinteni azokat a legfontosabb állomásokat, amelyek nélkül ez a munka nem születhetett volna meg.

Negyedéves hallgató koromban kezdetem TDK munkát végezni a Semmelweis Egyetem I. Gyermekklinikáján **Vásárhelyi Barna** vezetésével. Neki köszönöm, hogy a tudományos kutatás rendszerével megismertetett, megtanított tudományos közlemények és előadások készítésére és elindított kutatói pályámon. A TDK munka során együtt dolgoztam több hallgatótársammal, **Kocsis Istvánnal, Vannay Ádámmal, Tóth Judittal és Fekete Andreával**, akiknek köszönöm az akkori és az azóta eltelt idő alatt végzett kutatások során nyújtott segítségét.

Az egyetem elvégzése után PhD hallgató lettem **Szabó András** vezetése alatt. Neki köszönöm, hogy az experimentális kutatás módszertanával megismertetett, hogy a külföldi együttműködések fontosságára és a kutatás rendszerezettségének szükségességére felhívta figyelmemet. Köszönöm PhD hallgató társaimnak, elsősorban **Treszl Andrásnak és Krikovszky Dórának** a PhD során végzett kutatásokban nyújtott együttműködést.

A PhD képzés után a Berlini Charité Patológiai Intézetében kezdtem dolgozni ösztöndíjasként. Köszönöm mentoromnak, **Reinhold Schäfer**-nek, hogy a türelem és az empátia fontosságára felhívta figyelmemet és megismertette velem a német kutatási rendszer felépítését. Köszönöm ottani poszt doktor kollegáimnak, **Violetta Serrá-nak, Pawel Surowiak-nak és Jan Budczies-nak** az azóta is tartó együttműködést.

Visszatérve Magyarországra, újra a Semmelweis Egyetemen kezdtem dolgozni, és azóta is **Tulassay Tivadar** kutatócsoportjában vagyok. Neki köszönöm, hogy lehetővé tette a kutatásokat, hogy önzetlen támaszt nyújtott kritikus pillanatokban, valamint a megoldhatatlannak tűnő kérdésekhez adott tanácsait.

Az egyetemi kollegák közül köszönöm az elmúlt évek alatt nyújtott szakmai együttműködést **Szász Marcellnek, Tímár Józsefnek, Rásó Erzsébetnek, Dinya Eleknek, Haltrich Irénnek, Szőke Dominikának, Tulassay Zsoltnak, Kulka Janinának, Nagy Gyula Richárdnak, Rigó Jánosnak, Molnár Bélának és Galamb Orsolyának**. Köszönöm az együttműködés során nyújtott munkát a Semmelweis Egyetemen kívüli együttműködő hazai kutatási partnereimnek, **Pongor Sándornak, Szállási Zoltánnak, Pete Imrének, és Likó Istvánnak**.

Külön köszönöm PhD hallgatóimnak, **Munkácsy Gyöngyinek, Györffy Andrásnak, Tegze Bálintnak, Pénzváltó Zsófiának és Mihály Zsuzsannának** a közös munkát. Meg kell említeni azokat a graduális hallgatókat is, akik egyetemi diplomamunkájukat készítették témavezetésemmel: **Lánczky Andrást, Benke Zsombort, Kokas Mártont, Sztupinszki Zsófiát, Weltz Boglárkát, Tilesch Esztert, Vikor Attilát és Balázs Bálintot**.

Végül köszönöm a kutatásokat anyagilag lehetővé tevő szervezetek támogatását: **OTKA, Európai Unió, Kutatási és Technológiai Innovációs Alap, ETT, Magyar Ösztöndíj, Semmelweis Egyetem, Breast Cancer Research Foundation, Fuss a Rákkutatásért Alapítvány és az NFÜ**.